Machine Learning 1

Lecture 5.1 - Supervised Learning
Bayesian Linear Regression - The Equivalent Kernel

Erik Bekkers

(Bishop 3.3.3)

Image credit: Kirillm | Getty Images

Slide credits: Rianne van den Berg

Equivalent Kernel Formulation

- predictive distribution

$$p(t'|x',\mathbf{X},\mathbf{t},\alpha,\beta) = \int p(t'|x',\mathbf{w},\beta)p(\mathbf{w}|\mathbf{X},\mathbf{t},\alpha,\beta)\mathrm{d}\mathbf{w}$$

$$= \mathcal{N}(t'|\overbrace{\mathbf{m}_N^T\boldsymbol{\phi}(x')}^{y(x', m_N)}, \sigma_N^2(x'))$$

w_{MAP}

$$\overset{\in\mathbb{R}^M}{\mathbf{m}_N} = \beta\mathbf{S}_N\boldsymbol{\Phi}^T\mathbf{t} \qquad \sigma_N^2(x') = \frac{1}{\beta} + \boldsymbol{\phi}(x')^T\mathbf{S}_N\boldsymbol{\phi}(x') \qquad \mathbf{S}_N^{-1} = \alpha\mathbb{1} + \beta\boldsymbol{\Phi}^T\boldsymbol{\Phi}$$

$$\boldsymbol{\Phi} = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \dots & \phi_{M-1}(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_N) & \dots & \phi_{M-1}(\mathbf{x}_N) \end{pmatrix}$$

- predictive mean:

$$y(x', \mathbf{m}_N) = \boldsymbol{\phi}(x')^T\mathbf{m}_N$$

$$= \beta\, \underline{\phi}(x')^T S_N \overbrace{\underline{\Phi}^T \underline{t}} = \beta\, \underline{\phi}(x')^T S_n \overbrace{\sum_{n=1}^{N} \underline{\Phi}^T_{:n} t_n}^{\in\mathbb{R}^M}$$

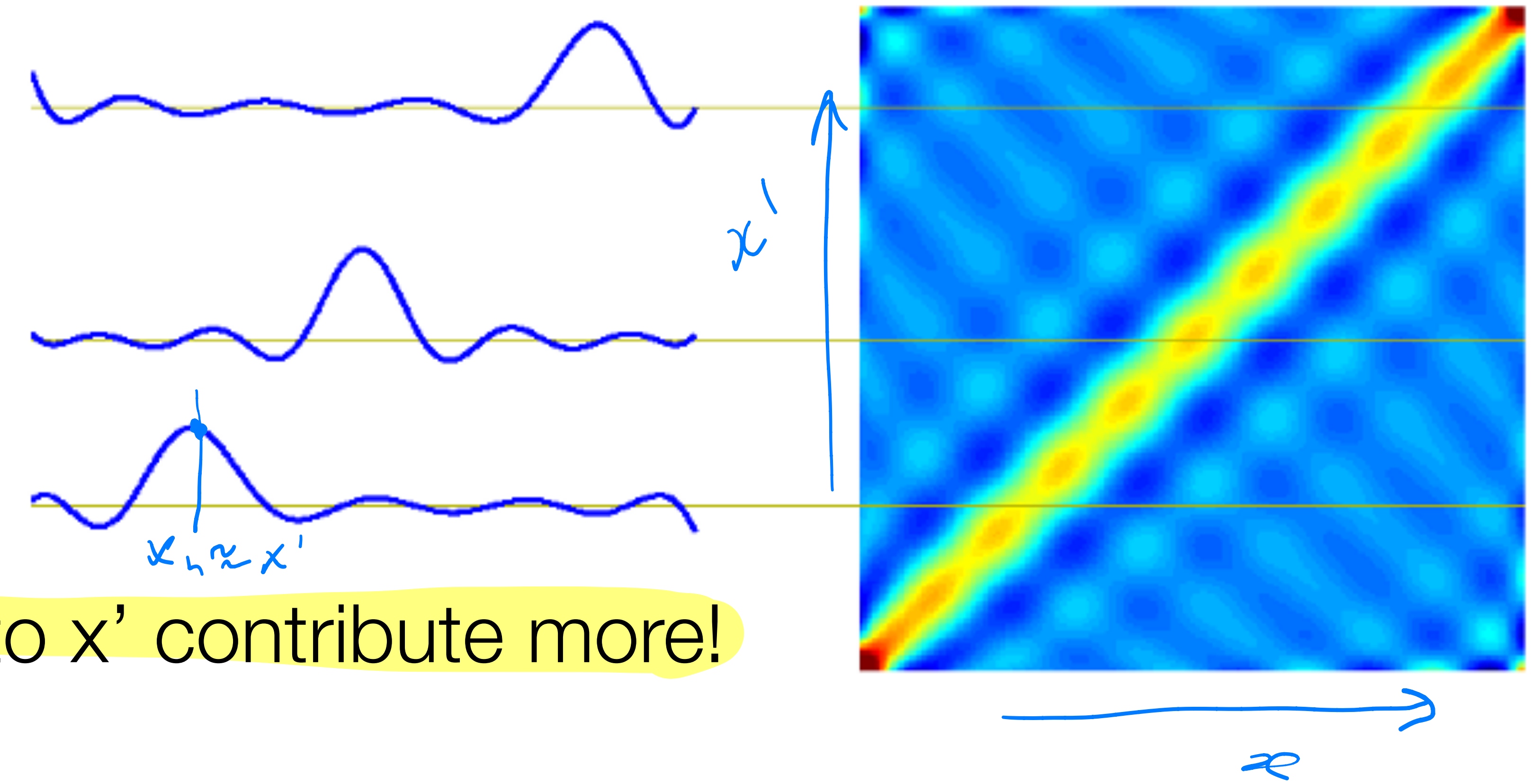
$$= \sum_{n=1}^{N} \underbrace{\beta\, \underline{\phi}(x')^T S_n\, \underline{\phi}(x_n)} t_n = \sum_{n=1}^{N} \underbrace{k(x', x_n)} t_n$$

- Equivalent kernel

$$k(x', x) = \beta\, \underline{\phi}(x')^T S_n\, \underline{\phi}(x_n)$$

Equivalent kernel for Gaussian Basis Functions

Figure: Equivalent kernel k(x’, x) (Bishop 3.10)



- Localized kernel

$$k(x', x) = \beta \boldsymbol{\phi}(x')^T \mathbf{S}_N \boldsymbol{\phi}(x)$$

- predictive mean

$$y(x', \mathbf{m}_N) = \sum_{n=1}^{N} k(x', x_n) t_n$$

- Training points x_n close to x’ contribute more!

- Covariance of between predictions:

$$\mathrm{cov}[t_1, t_2 | x_1, x_2] = \mathrm{cov}_{\mathbf{w}}[y(x_1, \mathbf{w}), y(x_2, \mathbf{w})]$$

$$= \mathrm{cov}_{\mathbf{w}}[\boldsymbol{\phi}(x_1)^T \mathbf{w}, \mathbf{w}^T \boldsymbol{\phi}(x_2)] = \mathbb{E}_w[\phi(x_1)^T w\, w^T \phi(x_2)] - \mathbb{E}_w[\phi(x_1)^T w]\, \mathbb{E}_w[w^T \phi(x_2)]$$

$$= \phi(x_1)^T \left(\mathbb{E}_w[w w^T] - \mathbb{E}_w[w]\, \mathbb{E}[w^T] \right) \phi(x_2) = \phi(x_1)^T \mathrm{cov}[w, w]\, \phi(x_2)$$

$$= \phi(x_1)^T S_N \phi(x_2) = \frac{1}{\beta} k(x_1, x_2)$$

$$p(\mathbf{w}|\mathbf{X}, \mathbf{t}, \alpha, \beta) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N)$$

$$\mathbb{E}[t'|x', \mathbf{w}] = y(x', \mathbf{w})$$